

Vanderbilt University Law School

Scholarship@Vanderbilt Law

Vanderbilt Law School Faculty Publications

Faculty Scholarship

2019

Policing, Databases, and Surveillance: Five Regulatory Categories

Christopher Slobogin

Vanderbilt University Law School

Follow this and additional works at: <https://scholarship.law.vanderbilt.edu/faculty-publications>



Part of the [Privacy Law Commons](#)

Recommended Citation

Christopher Slobogin, *Policing, Databases, and Surveillance: Five Regulatory Categories*, 13 Actual Problems of Economics & Law. 1093 (2019)

Available at: <https://scholarship.law.vanderbilt.edu/faculty-publications/1127>

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarship@Vanderbilt Law. It has been accepted for inclusion in Vanderbilt Law School Faculty Publications by an authorized administrator of Scholarship@Vanderbilt Law. For more information, please contact mark.j.williams@vanderbilt.edu.

POLICING, DATABASES, AND SURVEILLANCE: FIVE REGULATORY CATEGORIES

*Christopher Slobogin**

Databases are full of personal information that law enforcement might find useful. Government access to these databases can be divided into five categories: suspect-driven; profile-driven; event-driven; program-driven and volunteer-driven. This chapter recommends that, in addition to any restrictions imposed by the Fourth Amendment (which currently are minimal), each type of access should be subject to its own regulatory regime. Suspect-driven access should depend on justification proportionate to the intrusion. Profile-driven access should likewise abide by a proportionality principle but should also be subject to transparency, vetting, and universality restrictions. Event-driven access should be cabined by the time and place of the event. Program-driven access should be authorized by legislation and by regulations publicly arrived-at and evenly applied. Information maintained by institutional fiduciaries should not be volunteered unless necessary to forestall an ongoing or imminent serious wrong.

INTRODUCTION

It is now a commonplace that virtually everything we do is memorialized on databases, some of them maintained by government, some of them in the hands of private enterprises. These databases—which for ease of reference this chapter will refer to as *The Cloud*—reside in the servers of Google, Netflix and Apple; the memory banks of phones, closed-circuit cameras, “smart cars,” and satellites; and the computers in government agencies and commercial establishments. They track an astonishing range of our intimate daily activities, including Internet usage, communications connections, financial transactions, travel routes, tax information, medical treatment, and biometric information, as well as more prosaic matters such as employment and residence history, utility usage,

* Milton Underwood Professor of Law, Vanderbilt University. A version of this paper was published in the National Constitution Center’s White Paper Series, as *Policing and the Cloud*, available at constitutioncenter.org.

and car malfunctions. The question addressed here is when the government should be able to gain access to this wealth of personal information for law enforcement and national-security purposes.

In the United States, answering that question requires consulting a welter of statutes and a few Supreme Court decisions. For instance, when the government wants to access communications stored on a computer, federal and state laws usually require a warrant, issued by a judge who has found probable cause that the communication will lead to evidence of wrongdoing.¹ However, if officials want an opened message that is sitting on a server, or an unopened text that has been on a server for over 180 days, then they may only need to show that it is “relevant” to an investigation—a much lower standard than probable cause, albeit an assertion that at some point is challengeable by the target, as occurs with an ordinary subpoena.² And if the communication sits on a “private” server (belonging, say, to an employer), no court process is required.³

When law enforcement officials seek records from third parties outside the communications context, a wide array of statutes may be applicable. As a general matter, bank, educational, and even medical records can be obtained with a mere subpoena, which the target often does not find out about unless and until prosecution occurs.⁴ In a host of other situations, such as accessing commercial camera footage or obtaining data about credit-card purchases or past travel routes, most jurisdictions do not require police to follow any judicial process, but rather allow them to obtain the information at their discretion and that of the data holders.⁵ When law enforcement seeks information from the databases of other government agencies, as opposed to those maintained by private entities, usually all it needs is a written request from the head of the enforcement agency, although sometimes more is required.⁶

In theory, the U.S. Constitution, and in particular, the Fourth Amendment, could have something to say about all of this. The Fourth Amendment requires that the government act reasonably when it engages in a “search” or “seizure,” and the courts have held that, for many types of

¹ See, e.g., 18 U.S.C. §§ 2511 & 2518.

² 18 U.S.C. § 2703(a), (b)(1)(B). On February 9, 2017, the House of Representatives unanimously voted to repeal this provision and instead require a warrant; the Senate had yet to vote at the time of this writing.

³ 18 U.S.C. § 2711(2) (defining remote computing service).

⁴ For a summary, see CHRISTOPHER SLOBOGIN, *PRIVACY AT RISK: THE NEW GOVERNMENT SURVEILLANCE AND THE FOURTH AMENDMENT* 174–75 (2007).

⁵ See generally Erin Murphy, *The Politics of Privacy in the Criminal Justice System: Information Disclosure, the Fourth Amendment, and Statutory Law Enforcement Exemptions*, 111 MICH. L. REV. 485 (2013).

⁶ See SLOBOGIN, *supra* note 4, at 173 (describing the Privacy Act).

searches, this reasonableness requirement can only be met with a warrant. However, this requirement only applies to government actions that are considered “searches.” The Supreme Court has defined that word very narrowly, to encompass only those actions that infringe “reasonable expectations of privacy” or that involve some type of physical intrusion.⁷ Most relevant here are the Court’s decisions holding that expecting constitutional protection from government acquisition of information surrendered to third parties—whether they be Internet service providers, banks, or phone companies—is not reasonable, since we “assume the risk” that those third parties will decide to give that information to the government.⁸ As discussed below, this “third party” doctrine has seen some erosion in recent years, but it remains the reason that, other than when access to the content of communications is involved,⁹ the Fourth Amendment has had very little impact on the government’s ability to obtain information from private databases, even when it relies on technology to do so.

When instead the database is created by law enforcement, the Constitution may have more impact. In particular, collection of the information for the database may require justification. For instance, taking a DNA sample through a cheek swab is a Fourth Amendment search,¹⁰ and forcing an individual to produce self-incriminating documents can implicate the Fifth Amendment unless the government can identify relatively precisely the documents it wants.¹¹ However, any important regulatory need will overcome Fourth Amendment claims that these types of data acquisition are unreasonable; in such cases, probable cause is not necessary.¹² And if the information is “nontestimonial” (as is assumed to be the case with fingerprints and DNA¹³), or is “voluntarily” surrendered for non-criminal purposes (as is assumed to be the case with a tax return or

⁷ *Florida v. Jardines*, 133 S. Ct. 1409, 1414 (2013) (indicating that the expectation of privacy test established in earlier case law is supplemented by inquiry into whether the government “engage[s] in [a] physical intrusion of a constitutionally protected area”).

⁸ *See, e.g., United States v. Miller*, 425 U.S. 435, 443 (1976) (holding one has no expectation of privacy in bank records, “even if the information is revealed on the assumption that it will be used only for a limited purpose and the confidence placed in the third party will not be betrayed”); *Smith v. Maryland*, 442 U.S. 735, 743–44 (1979) (same holding with respect to phone numbers dialed).

⁹ *See, e.g., United States v. Warshak*, 631 F.3d 266 (6th Cir. 2010) (holding that the Fourth Amendment requires a warrant to obtain stored emails).

¹⁰ *See Maryland v. King*, 133 S. Ct. 1958, 1969–70 (2013).

¹¹ *See United States v. Hubbell*, 530 U.S. 27, 45 (2000).

¹² *See, e.g., King*, 133 S. Ct. at 1977 (holding that the government’s need for DNA from arrestees outweighs the intrusion involved).

¹³ *See, e.g., Wilson v. Collins*, 517 F.3d 421, 431 (6th Cir. 2008).

applications for government entitlements¹⁴), or is obtained from a source other than individual, the Fifth Amendment doesn't apply. Finally, the Constitution appears to have little to say about law enforcement agencies' access to the information once they or other government entities legitimately collect it.¹⁵

While many have criticized the laxness of both statutory and constitutional law, the most popular counter-proposal—that all or most Cloud access by the government should require a judicial warrant—has problems of its own. Conceptually, a warrant requirement glosses over the intuition that a large number of situations, while involving a viable privacy claim against the government, do not merit the full protection of a judicial probable-cause finding. Practically, it would handcuff legitimate government efforts to nab terrorists and criminals. A more nuanced approach is probably necessary.

That approach should begin with an assessment of the varying motivations that drive the government's use of The Cloud. Cloud-based searches can come in at least five different guises: suspect-driven, profile-driven, event-driven, program-driven, or volunteer-driven. Some database access by the state is aimed at getting as much information as possible about individuals suspected of wrongdoing. Other efforts do not start with a particular suspect, but rather with a profile of a hypothetical suspect, purportedly depicting the characteristics of those who have committed or will commit a particular sort of crime. A third type of Cloud-search starts neither with a suspect nor a suspect profile but with an event—usually a crime—and tries to figure out, through location and related information, who might be involved. Fourth, so as to have the information needed for suspect-, profile-, and event-driven operations at the ready, government might initiate data-collection programs. Finally, the government also relies on citizens to come forward on their accord when they find incriminating information about another person in The Cloud.

¹⁴ See, e.g., *Garner v. United States*, 424 U.S. 648, 657–58 (1976) (holding that the federal penalty for failing to file a tax return does not coerce answers to individual questions on the return, which the taxpayer can answer by asserting the privilege with impunity); *Balt. Dep't of Soc. Serv. v. Bouknight*, 493 U.S. 549, 556 (1990) (stating that “the Fifth Amendment privilege may not be invoked to resist compliance with a regulatory regime constructed to effect the State's public purposes unrelated to the enforcement of its criminal laws”).

¹⁵ See Erin Murphy, *DNA in the Criminal Justice System: A Congressional Research Service Report*, 64 UCLA L. REV. DISCOURSE 340, 364 (2016) (noting that even with respect to accessing genetic databases that can contain extremely personal information, “[s]tandards surrounding the legality of both [on-demand and volunteered] disclosure have not yet been fully adjudicated in the courts”).

Each of these endeavors is distinct from the other four. Each calls for a different regulatory regime. Below is a sketch of what those regimes might look like. While they borrow from Fourth Amendment jurisprudence, the principles developed here fill a void because, to date, that jurisprudence has had little to say about Cloud searches. Until the courts weigh in more definitively, policymakers are working pretty much on a clean slate in this area.

I. SUSPECT-DRIVEN CLOUD ACCESS-PROPORTIONALITY

Assume the police receive an anonymous phone call from a female claiming that John Slade, a fifth-grade public-school teacher, is also a drug dealer. In investigating this claim, police might want to obtain Slade's phone records to see if he has called known drug dealers, gang members, or drug users. They might also seek access to his bank records to discover whether the amount of money he deposits is consistent with his job as a school teacher. Additionally, the police might like to find out from GPS records and drone and camera feeds if Slade frequents areas of town where drugs are routinely sold.

Under current Fourth Amendment and statutory law, none of these policing moves requires a warrant or probable cause and, depending on the jurisdiction, some of them may not even require a subpoena. That lack of regulation is abetted by the Supreme Court's assertion that expecting privacy in information surrendered to a third party or in activities carried out in public is unreasonable.¹⁶ Yet most people surveyed on these matters come to a quite different conclusion, ranking perusal of their bank and phone records, for instance, as comparable to search of a bedroom, and ranking location tracking as similar in invasiveness to a frisk.¹⁷ On a more philosophical plane, scholars argue that allowing the government to invade databases so easily offends not only privacy, but autonomy and dignity.¹⁸

¹⁶ See *Jardines*, 133 S. Ct. at 1414; *United States v. Knotts*, 460 U.S. 276, 281 (1983) (no expectation of privacy in public).

¹⁷ See SLOBOGIN, *supra* note 4, at 112 tbl. (compare items 14 and 16), 184 tbl. (compare items 24 and 25). This research has been replicated in Christine S. Scott-Hayward, Henry F. Fradella & Ryan G. Fischer, *Does Privacy Require Secrecy?: Societal Expectations of Privacy in the Digital Age*, 43 AM. J. CRIM. L. 19 (2015); and Jeremy E. Blumenthal, Meera Adya & Jacqueline Mogle, *The Multiple Dimensions of Privacy: Testing "Lay" Expectations of Privacy*, 11 U. PA. J. CONST. L. 331, 345 (2009).

¹⁸ See, e.g., David Lametti, *The Cloud: Boundless Digital Potential or Enclosure 3.0?*, 17 VA. J.L. & TECH. 190, 196 (2012) ("[W]e may be witnessing another round of 'enclosure' in Cloud space that might have serious deleterious effects on what we have come to expect in the digital age: autonomy, exchange, spontaneity, and creativity, and all at a lightning pace.").

They also claim it chills citizens' rights to expression and association, and creates huge potential for abuse; after all, knowledge—which The Cloud provides in troves—is power.¹⁹

The Supreme Court itself has begun to recognize these concerns. In *Riley v. California*,²⁰ despite centuries-old precedent permitting suspicionless searches of any item found on an arrested individual, it required a warrant for a search of a cell phone of an arrestee, in recognition of the fact that “the cell phone collects in one place many distinct types of information—an address, a note, a prescription, a bank statement, a video—that reveal much more in combination than any isolated record.”²¹ In *United States v. Jones*,²² five members of the Court concluded that a Fourth Amendment search occurs when the police engage in “prolonged” tracking of a vehicle using GPS signals. While neither *Riley* nor *Jones* involved database access, Justice Sotomayor may have summed up where the Court is going when she stated in her concurring opinion in *Jones* that “it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties. This approach is ill-suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks.”²³

On this view, government would not be foreclosed from perusing, at its discretion, blogs, tweets, public records, and other sources that are clearly meant to be consumed by the public. But it would prohibit police from accessing, in the absence of justification, non-public Cloud data people generate when they engage in “mundane tasks” like communicating with their friends, banking, and shopping. It would also prohibit suspicionless access to tracking data about everyday travels that the average person undertakes on the assumption of practical anonymity.

In short, there is a strong case to be made for requiring the government to demonstrate it has good reason to go after Cloud-based information about a particular person that is not readily available in public fora. Then the question becomes how good that reason must be. Normally, the Fourth Amendment requires that a search be based on probable cause, which amounts to a “fair probability” that a search will discover evidence of

¹⁹ See, e.g., Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1461 (2001) (“The problem with databases is ... a problem that involves power and the effects of our relationship with public and private bureaucracy—our inability to participate meaningfully in the collection and use of our personal information.”).

²⁰ 134 S. Ct. 2473 (2014).

²¹ *Id.* at 2489.

²² 565 U.S. 400 (2012).

²³ *Id.* at 417 (Sotomayor, J., concurring).

crime.²⁴ Return to the investigation of Slade. If the caller had identified herself and provided detail about Slade's drug deals, perhaps the police would have had probable cause and grounds for a full-scale digital search. But recall that, in fact, the caller was anonymous and simply said Slade was selling drugs, thus making it difficult to dismiss the possibility that she was a disgruntled student or a spurned lover. Under Supreme Court case law, that call, by itself, would not permit a traditional search.²⁵

But suppose instead that the call, although anonymous, provides detail about John's next drug deal. While, by itself, this would not be enough for probable cause, its predictive quality does provide an additional indication of reliability.²⁶ In that intermediate situation, police arguably have "reasonable suspicion" (a lesser level of cause but one that nonetheless requires an articulable reason to act).²⁷ In that scenario, police might still be prohibited from requisitioning the capacious digital record described above. But perhaps they would be justified in seeking more limited transactional data, say information about whether, near the predicted time, Slade calls a particular number or heads toward a particular location.

This measured approach to accessing The Cloud is based on what might be called the proportionality principle.²⁸ Under traditional Fourth Amendment rules, an arrest requires probable cause, but a short detention only reasonable suspicion; similarly, a full search of the person requires probable cause, a frisk only reasonable suspicion.²⁹ Analogously, significant invasions of privacy on The Cloud—obtaining a month's worth of bank records or Internet logs, or as the Supreme Court itself suggested in *Jones*, travel records that track a person for four weeks³⁰—might require cause about the target akin to that necessary to search a home or car. However, less significant invasions—accessing records about a single phone call, credit-card purchase, or car trip, pulling up an identity using facial-recognition technology, or tracking a car for a few hours—could be justifiable on something less. Not only does this type of proportionality principle better reflect the degree of the government's intrusion, it also avoids the Catch-22 of requiring police to demonstrate probable cause

²⁴ See WAYNE R. LAFAYE ET AL., 2 CRIMINAL PROCEDURE 114–15 (3d ed. 2007).

²⁵ See *Florida v. J.L.*, 529 U.S. 266 (2000) (holding unconstitutional a frisk based on an anonymous phone call stating that the defendant would be standing on a street corner wearing certain clothing with a gun on his person).

²⁶ Cf. *Illinois v. Gates*, 462 U.S. 213 (1983) (holding police had probable cause based on an anonymous letter that provided considerable predictive detail, but only after some of the detail was corroborated by police).

²⁷ See *Terry v. Ohio*, 392 U.S. 1, 27 (1968).

²⁸ See SLOBOGIN, *supra* note 4, ch. 2.

²⁹ See *Terry*, 392 U.S. at 20–27.

³⁰ *Jones*, 565 U.S. at 403.

before carrying out the preliminary investigative techniques they need to develop it.

Proportionality reasoning makes sense in the abstract. But it presents difficult line-drawing problems. What justification do police need if, rather than seeking data about Slade's financial transactions or travels over the course of a month, they want only a week's worth of data? Or if they want to ascertain, in combination, whether Slade calls a particular number, visits a particular location, and deposits a large amount of money during a given month, but seek no other information about him?

Answers to these types of questions will inevitably produce somewhat arbitrary classifications. Sometimes the answer might be categorical. That was the angle the Supreme Court took with respect to searches of home interiors carried out with sophisticated technology; in *Kyllo v. United States*, the Court held that *all* such searches require probable cause.³¹ Government access to Cloud data that is analogous to the interior of the home—for instance, private documents stored on The Cloud, or communications on a closed social network—should receive similar categorical protection.³²

Once data leaves such confines, however, an across-the-board warrant requirement for accessing personal information overprotects the interests at stake, as both the Court's cases and people's views on the matter suggest.³³ One approach is to differentiate between types of information. Perhaps medical records would receive the most protection, bank records something less, utility records something less still.³⁴ Current federal law appears to adopt this approach with respect to communications, with subscriber information receiving minimal protection, phone numbers and e-mail addresses receiving more protection, stored communications even more,

³¹ 533 U.S. 27, 37–38 (2001) (involving thermal imaging of the home).

³² Some have argued that encrypted material should receive similar, or even *absolute*, protection, simply by virtue of being encrypted. But given the fact that anything, including impersonal business records, can be encrypted, proportionality reasoning would suggest that the government should be able to force decryption of any material for which it has the requisite cause. The encryption debate is too complicated to address in this limited space. See Hugh J. McCarthy, *Decoding the Decryption Debate: Why Legislating to Restrict Strong Encryption Will Not Resolve the "Going Dark" Problem*, 20 J. INTERNET L. 1 (2016) (detailing practical problems and domestic and international legal issues associated with different approaches designed to permit government decryption).

³³ See Christopher Slobogin, *Government Data Mining and the Fourth Amendment*, 75 U. CHI. L. REV. 317, 335 (2008) (presenting survey data indicating significantly different "intrusiveness ratings" depending on the type of record accessed).

³⁴ For an effort in this vein, see AM. BAR ASS'N, ABA STANDARDS FOR CRIMINAL JUSTICE: LAW ENFORCEMENT ACCESS TO THIRD PARTY RECORDS (3d ed. 2013), https://www.americanbar.org/content/dam/aba/publications/criminal_justice_standards/third_party_access.authcheckdam.pdf.

and interception of communications requiring probable cause.³⁵ But the intuition upon which this scheme is based is suspect: For instance, a month's worth of "metadata" about a person's contacts may reveal much more than the transcript of a conversation.³⁶ Similar comments can be made about other types of data: Bank records, credit-card statements, and utility logs can all be more or less private depending on the person and the context.

In these circumstances, an alternative or supplemental proportionality approach might rely on durational or aggregational limitations. In *Jones*, five members of the Court distinguished between "short-term" and "prolonged" tracking.³⁷ Similarly, the Court has indicated that, while a physical seizure lasting less than 15 minutes usually requires reasonable suspicion, a longer seizure amounts to an arrest requiring probable cause,³⁸ and an arrest must be judicially reviewed within 48 hours.³⁹ One might limit Cloud searches of non-public data outside the home context the same way, on the theory that the more one learns about a person—from whatever source—the more intrusion occurs. For instance, obtaining information about the transactions of someone like Slade on a particular day or over a couple of days might be permitted on a relevance showing, but seeking data shadowing his activities over more than a 48-hour period or with respect to several different days might require greater suspicion and a subpoena from a judge, and obtaining a month's worth of transactions could require probable cause and a warrant. While this duration-based rule also has administrability problems,⁴⁰ it has the benefit of simultaneously protecting privacy in a roughly proportionate manner and permitting government to build its case without requiring probable cause from the outset. Ultimately, policymakers applying proportionality reasoning to suspect-driven Cloud access might choose rules based on a combination of record-type and aggregation considerations.

Even if one finds this type of reasoning persuasive in the abstract, it might be resisted in the specific context of national-security investigations.

³⁵ See *supra* notes 1–4 (citing relevant statutes).

³⁶ See Steven M. Bellovin et al., *It's Too Complicated: How the Internet Upends Katz, Smith, and Electronic Surveillance Law*, 30 HARV. J. L. & TECH. 1, 92 (2016) (given technological developments, "[t]he concept of metadata as a category of information that is wholly distinguishable from communications content and thus deserving of lower privacy protection is no longer tenable").

³⁷ *Jones*, 565 U.S. at 430–31 (Alito, J., concurring).

³⁸ See generally *United States v. Sharpe*, 470 U.S. 675, 684–88 (1985).

³⁹ *Cty. of Riverside v. McLaughlin*, 500 U.S. 44, 56 (1991).

⁴⁰ Compare Orin Kerr, *The Mosaic Theory of the Fourth Amendment*, 111 MICH. L. REV. 311 (2011) (describing some of the problems); with Christopher Slobogin, *Making the Most of United States v. Jones in a Surveillance Society: A Statutory Implementation of Mosaic Theory*, 8 DUKE J. CONST. L. & PUB. POL'Y 1, 24–30 (2012) (providing a model statute implementing mosaic theory).

Where national security is at stake, the argument goes, any significant limitation on Cloud usage should be jettisoned. But this stance should be viewed with skepticism. “National security” is an extremely broad term, and it has too often been a blank check for government abuse.⁴¹ Concrete threats to the country might justify departure from the rules that normally govern domestic law enforcement; for instance, if there is a demonstrable, significant, and imminent danger, relaxation of the justification required by proportionality reasoning might be permissible in this context. But otherwise the National Security Agency and like government entities should probably be treated no differently than other law enforcement agencies.

II. PROFILE-DRIVEN CLOUD ACCESS-HIT RATES

Profile-driven searches are very similar to suspect-driven searches. The difference is that suspect-driven searches start with a person thought to be engaged in wrongdoing and then go to The Cloud, while with profile-driven searches the government has no particular suspect when it seeks out Cloud data; rather it utilizes a profile describing the characteristics of likely perpetrators that it hopes will identify wrongdoers. Again using John Slade as an example, imagine that the police focus on him not because of an anonymous tip but because of a drug-dealer profile developed with the help of computer scientists and criminologists. Such a profile might be composed, let’s say, of five factors having to do with travel, spending, and communication patterns. Or, similar to how credit-card companies identify theft and fraud, the profile might purport to tell police when and where a drug deal is occurring or is soon likely to occur, which allows them to conduct surveillance of that spot and perhaps nab a perpetrator. Analogous to how researchers have developed risk-assessment instruments for pretrial detention and sentencing purposes,⁴² these profiles would initially be based on analysis of drug-dealer characteristics and behavior, and then cross-validated on new populations or locations.

Profiling using data accumulated from Cloud-related sources, sometimes called “predictive policing,” is in its infancy. But police departments appear to be committed to developing the necessary tools.⁴³

⁴¹ See generally Erik K. Yamamoto, *White (House) Lies: Why the Public Must Compel the Courts to Hold the President Accountable for National Security Abuses*, 68 LAW & CONTEMP. PROBS. 285 (2005).

⁴² See e.g., Christopher T. Lowenkamp & Jay Whetzel, *The Development of an Actuarial Risk Assessment Instrument for U.S. Pretrial Services*, 73 FED. PROBATION 33 (Sept. 2009).

⁴³ See Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 352–88 (2015); Jeffrey Fagan, “Race and the New Policing,” in the present Report.

Such profiles are only useful, of course, if the government has access to databases that have the information needed to run the profile. Whether it should have such access is discussed below (under program-driven Cloud searches). Assume for now the data is available to government officials.

As with suspect-driven Cloud searches, the analysis of profile-driven Cloud inquiries should involve determining whether the justification is proportional to the intrusion. In other words, the profile must produce a “hit rate” equivalent to the certainty required by the proportionality principle. If one equates probable cause with approximately a 50% hit rate, a profile that correctly identifies a drug dealer only 20% of the time would not authorize use of a profile that accesses multiple intimate data sources. But it might justify use of a profile that relies on arrest records, gang member lists, and other public or quasi-public data.

Achieving even a 20% rate may be impossible for most crime scenarios, however; certainly social scientists engaged in the analogous pursuit of predicting dangerousness for sentencing purposes have struggled to achieve such accuracy. There are scores of variables associated with criminal behavior, and the prognostic power of any given variable or combination of variables is likely to be very low. Further, profiles will probably need to be updated routinely, either because of naturally occurring changes in criminal behavior or because perpetrators get wind of the factors in the profile. When one adds to those challenges the fact that much of the information about individuals found on The Cloud is unreliable,⁴⁴ profiles that might justify apprehending specific suspects will be few and far between, at least if police action based on such data abides by the proportionality principle.

Assuming that profiles with acceptable hit rates can nonetheless be developed, a second limitation on profile-driven Cloud use is that it should be transparent. To avoid profiles concocted after the fact, allow perusal of hit-rate data, and ensure that those individuals who are targeted using a profile actually meet it, profiles must be accessible to courts and other oversight entities, at least on an *in camera* basis (i.e., in chambers, outside the presence of the public). Transparency also assures that the factors on which profilers rely are vetted to ensure that illegitimate ones, such as those that are racially discriminatory, are not influencing the results.

This vetting process could become difficult if, as occurs in some commercial contexts, profiles rely on complex algorithms generated through opaque machine-learning techniques or protected from disclosure for proprietary reasons.⁴⁵ Complicating matters further, risk factors such as

⁴⁴ See Ferguson, *supra* note 43, at 398–99.

⁴⁵ See Michael L. Rich, *Machine Learning, Automated Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 883–86 (2016).

criminal history, location, and employment may turn out to be proxies for race, class, and related traits, use of which are generally considered anathema in police work.⁴⁶

These concerns do not have to be paralyzing, however. For instance, profiles that are indecipherable could be banned in the law enforcement context, regardless of their accuracy,⁴⁷ or can be designed to ensure “procedural regularity.”⁴⁸ Steps can also be taken to alleviate the concern that some risk factors correlate with race as well as crime. For instance, developers of algorithms designed to detect potential hot spots or perpetrators could be directed to avoid arrest records for low-level or drug crimes that might reflect race-based policing practices; instead, developers can be told to rely on *reports* of crimes (for hot-spot profiles) and on crimes of violence or on property crimes (for suspect profiles), so as to reduce the influence of racially-discriminatory arrest rates for drug crimes and similarly bias-susceptible offenses.⁴⁹ It is also important to remember that traditional policing often relies on the same suspect, static factors, in ways that are inevitably more intuitive, and therefore less discoverable and more subject to invidious manipulation. Transparent algorithms that can produce the relevant hit rates and that avoid obviously illegitimate variables are very likely to be an improvement.⁵⁰

To limit further the extent to which bias creeps into the process, however, a third limitation that should be imposed on profile-driven Cloud searches is the maxim that everyone who fits a given profile must be treated the same. That means if a drug-dealer profile with the relevant hit rate

⁴⁶ See Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CAL. L. REV. 671 (2016). For discussions of race and policing, see David A. Harris, “Racial Profiling,” in the present Report; and Devon W. Carbado, “Race and the Fourth Amendment,” in the present Report.

⁴⁷ See Andrew D. Selbst & Solon Barocas, *Regulating Inscrutable Systems* (2017) (unpublished manuscript) (on file with authors) (identifying increasingly difficult-to-interpret approaches to algorithms, beginning with “decision tree” logic and ending with “deep learning” artificial intelligence).

⁴⁸ See Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017) (sketching how computer programs can be constructed to ensure fairness and procedural regularity despite the black box nature of machine learning).

⁴⁹ Cf. Michael Feldman et al., *Certifying and Removing Disparate Impact*, PROCEEDINGS OF THE 21ST ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING 259 (2015) (discussing similar moves in connection with hiring algorithms).

⁵⁰ See, e.g., Sharad Goel et al., *Combatting Police Discrimination in the Age of Big Data*, 20 NEW CRIM. L. REV. 181 (2017) (using stop and frisk data from New York City to create a risk profile that predicted who would be carrying a weapon 83% of the time; also finding that factors like “furtive movement,” a common police justification for stops, was not related to weapon possession and that, of those stopped using the profile, whites were much more likely than blacks to have a weapon).

identifies 200 people, police should not be able simply to single out someone like Slade but rather would either have to investigate everyone who fits the profile or, if that is not feasible, select individuals on a neutral, pre-specified basis (e.g., every third person). In the absence of this limitation, attempts to avoid illegitimate discrimination in construction of the profile will merely reappear at the post-profile investigation stage.

The added advantage of this third limitation on profile-driven actions is that it would make law enforcement think twice before engaging in them. Profile-driven searches will produce a large number of false positives, no matter how good they are. If, for instance, the predicted hit rate is 50%, half of those investigated are likely to be innocent, whether the police go after everyone identified by the algorithm or only a neutrally selected subgroup. Even if the post-profile police work is covert, much investigative energy will be expended with no gain. And in those situations where the investigation of those who meet the profile involves overt searching or seizing, a non-trivial number of false positives are likely to complain. Although the quantified, objective nature of profile-driven Cloud searches offers many advantages over traditional suspect-based techniques, their dragnet nature may end up being so practically or politically unpalatable when used to identify “persons of interest” that police abandon them.

III. EVENT-DRIVEN CLOUD ACCESS-HASSLE RATES

Some Cloud searches conducted by law enforcement start not with a suspect or a profile of a likely suspect, but with an event—usually a crime—and use Cloud data to try to figure out who perpetrated or witnessed it. Let’s return to the example of John Slade, but this time as a victim rather than a potential suspect. Imagine that at 2 a.m. one Sunday morning, police are called to the scene of a homicide, a dark urban street, where they find Slade dead, drugs strewn around him. A medical examiner says the death probably occurred two hours earlier, around midnight. Pre-Cloud, the police would probably go door to door talking to those who live in the immediate vicinity, some or all of whom might claim—honestly or not—to have been elsewhere at the relevant time or to have seen or heard nothing. In contrast, today police might access phone or vehicle GPS records, as well as feeds from surveillance cameras with face-recognition or night-vision capacity, to identify people or cars near the crime scene at the time it happened, and then use suspect-driven techniques to zero in on the perpetrator.⁵¹

⁵¹ Baltimore has used videos from plane cameras to “TiVo” backward from the scene of the crime to determine how individuals and vehicles got there. See Monte Reel, *Secret Cameras Record Baltimore’s Every Move from Above*, BLOOMBERG BUSINESSWEEK (Aug. 23, 2016), <https://www.bloomberg.com/features/2016-baltimore-secret-surveillance/>.

These event-driven uses of The Cloud could result in a large haul of people, among whom may be the perpetrator or a witness, but many of whom will be neither. At the same time, all that this “data dump” learns about any of these individuals is that they were near a particular place at a particular time, a discovery that proportionality reasoning would suggest requires little justification. Even so, the scope of the government’s Cloud inquiry should probably be limited, to reduce both the extent of the initial privacy invasion and the number of people subject to further law enforcement inquiry. In other words, the government should minimize what Jane Bambauer calls the “hassle rate”—the proportion of innocent people subject to police investigation in an effort to find the one or two bad people.⁵²

What that rate should be will depend on the likely number of people involved. In effect, an admonition to limit hassle rates is simply a call to shape event-driven searches around the relevant time and place. In investigating Slade’s death, for instance, police should be able to find out the identity of and question pedestrians and car drivers near the scene of the crime shortly before or after midnight (assuming the medical examiner’s assessment is correct). But perhaps the police should not be able to investigate people who never approached the scene closer than 50 yards or who were there before 11:30 p.m. or after 12:30 a.m.

The Cloud facilitates immensely the ability of investigators to carry out event-driven inquiries. Such inquiries can be quite broad, limited only by the imagination and priorities of law enforcement (because they are not limited by current law, at least in most jurisdictions). In contrast to the hit rates required for profile-driven Cloud searches, acceptable hassle rates for event-driven Cloud searches are not easy to establish, and should probably vary with the type of information sought and the type of crime being investigated.⁵³ If the law is called into play here, perhaps the best that can be done is to require police to seek authorization for such inquiries from a judge, who can take potential hassle rates and these other factors into account in determining whether and to what extent event-driven Cloud searches may occur.

⁵² Jane Bambauer, *Hassle*, 113 MICH. L. REV. 461 (2015).

⁵³ In an analogous situation, the Supreme Court held that the analysis should consider “the gravity of the public concerns served by the seizure, the degree to which the seizure advances the public interest, and the severity of the interference with individual liberty.” *Illinois v. Lidster*, 540 U.S. 419, 427 (2004) (upholding a roadblock at the time of day and the place of a hit-and-run accident committed one week earlier, set up to find possible witnesses).

IV. PROGRAM-DRIVEN CLOUD ACCESS-DEMOCRATIC AUTHORIZATION

Suspect-driven, profile-driven, and event-driven Cloud searches all rely in varying degrees on access to multiple databases, ranging from those that keep track of communications and travels to those that house records of financial and social transactions. From law enforcement's perspective, keeping these databases within their separate silos is, at the least, inefficient and, in the case of profile-driven Cloud access, perhaps fatal, since profiles usually only work when they can access several databases at once. It was in recognition of this fact that the Defense Department proposed, post-9/11, the Total Information Awareness (TIA) program. According to a chart prepared by the Department of Defense, TIA was meant to gather in one place a huge array of transaction information concerning, according to the official description, "financial, educational, medical, veterinary[!], entry [i.e., immigration and customs], transportation, housing, ... and communications" activities, as well as all government records.⁵⁴ Once collected, these data would be combed using algorithms designed to detect terrorist activity. Congress, apparently not enamored of this idea, defunded TIA in 2003 (by voice vote).⁵⁵ But if Edward Snowden is to be believed, several programs in operation today, run by the NSA or other government agencies, bear at least some resemblance to it.⁵⁶

As the public reaction to Snowden's revelations indicates, a significant proportion of the citizenry is uncomfortable with these types of programs. Compilation of information from multiple sources in one "place" raises a host of concerns. As recent exposés of foreign machinations highlight, aggregation of data facilitates hacking and identity theft.⁵⁷ It also leads to "mission creep," as law enforcement realizes that information obtained for one reason (such as fighting terrorism) might be useful for other purposes. It can easily lead to more obvious abuses, ranging from illegitimate investigations of journalists, politicians, activists, and members of certain

⁵⁴ See *Total Information Awareness*, WIKIPEDIA, https://en.wikipedia.org/wiki/Total_Information_Awareness (last updated July 14, 2017) (depicting a chart purporting to have been prepared by the Defense Advanced Research Projects Agency).

⁵⁵ See 149 Cong. Rec. S1379-02, S1416 (daily ed. Jan. 23, 2003).

⁵⁶ See Glenn Greenwald, *XKeyscore: NSA Tool Collects "Nearly Everything a User Does on the Internet,"* THE GUARDIAN (July 31, 2013), <https://www.theguardian.com/world/2013/jul/31/nsa-top-secret-program-online-data>.

⁵⁷ See, e.g., Nicole Perlroth & David Gelles, *Russian Hackers Amass over a Billion Internet Passwords*, N.Y. TIMES (Aug. 5, 2014), <https://www.nytimes.com/2014/08/06/technology/russian-gang-said-to-amass-more-than-a-billion-stolen-internet-credentials.html>.

ethnic groups to leaks based on personal vendettas.⁵⁸ Most prominently, it tempts the government to combine all of the information it has collected to create “personality mosaics” or “digital dossiers” about each of its citizens, a phenomenon classically associated with totalitarian states.⁵⁹

In part because of the public reaction to Snowden’s disclosures, the NSA supposedly no longer collects metadata and must now seek it through subpoenas from the relevant common carriers, in the suspect- and profile-driven manner described earlier.⁶⁰ But the NSA and other federal agencies continue to aggregate other types of data.⁶¹ Localities and states also engage in the data-collection enterprise. For instance, New York City’s Domain Awareness system, co-created by the city’s police department and Microsoft, collates information gleaned from thousands of closed-circuit surveillance cameras (CCTV), and combines it with geospatial data that reveals crime “hot spots,” feeds from license-recognition systems, and GPS signals that permit real-time and historical tracking of cars.⁶² A number of other cities operate large-scale CCTV systems, and many are also moving toward 24/7 drone or plane surveillance.⁶³ A different type of program, known as a “fusion center,” exists in more than half the states. These centers—over 75 at last count, some with more than 200 personnel—“fuse” financial, rental, utility, vehicular, and communications data from federal, state, and local public databases, law enforcement files, and private company records for investigative purposes.⁶⁴

⁵⁸ For some examples involving activists, journalists, members of minority groups and so on, see Robert H. Sloan & Richard Warner, *The Self, the Stasi, and the NSA: Privacy, Knowledge, and Complicity in the Surveillance State*, 17 MINN. J.L. SCI. TECH. 347, 347–80 (2016).

⁵⁹ Daniel Solove popularized the term “digital dossiers,” which he described as the aggregation of data to create “a profile of an individual’s finances, health, psychology, beliefs, politics, interests, and lifestyle” that “increasingly flows from the private sector to the government, particularly for law enforcement use.” Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1084 (2004).

⁶⁰ USA FREEDOM Act of 2015, Pub. L. No. 114-23, § 101, 129 Stat. 268, 269-71 (2015).

⁶¹ Zack Whittaker, *Freedom Act Will Kill Only One of NSA’s Programs (and Not Even One of Its Worst)*, ZERO DAY (May 4, 2014), <http://www.zdnet.com/article/freedom-act-metadata-phone-records-prism/#1>.

⁶² See Colleen Long, *NYPD, Microsoft Create Crime-Fighting “Domain Awareness” Tech System*, HUFFINGTON POST (Feb. 25, 2013).

⁶³ See Somini Sengupta, *Privacy Fears Grow as Cities Increase Surveillance*, N.Y. TIMES (Oct. 13, 2013); Marc J. Blitz et al., *Regulating Drones under the First and Fourth Amendments*, 57 WM. & MARY L. REV. 49 (2015).

⁶⁴ See THE CONST. PROJECT, RECOMMENDATIONS FOR FUSION CENTERS: PRESERVING PRIVACY AND CIVIL LIBERTIES WHILE PROTECTING AGAINST CRIME AND TERRORISM 4 (2012), www.constitutionproject.org/pdf/fusioncenterreport.

These program-driven efforts, which have been called “panvasive” because they invade the records of large swaths of the population, occur with the foreknowledge that most of those affected have done nothing wrong.⁶⁵ Thus, this collection of data cannot be regulated through suspicion-based proportionality reasoning. Arguably, however, it does not need to be. Until the data are accessed by humans and used as a means of investigating or identifying particular people like Slade, no concrete intrusion has occurred. Only when such access does occur will government officials need to demonstrate the cause necessary to carry out suspect-, profile-, or event-driven searches.

For those who do not trust government to abide by such strictures, one further protection, illustrated by Congress’ changes to the NSA’s metadata program, would be to require that all databases be maintained outside the government. Even profile-driven Cloud searches could be carried out by a private entity, with the government providing the profile and the company providing the government only with the identities of those who meet it. While this arrangement would still present some of the problems associated with aggregation (hacking and the like), it would undoubtedly reduce the potential for mischief by government officials.

In the end, however, this attempt to separate government from data cannot work. Many of the databases useful to Cloud searches—those that house CCTV feeds, the data from highway tracking systems, and the billions of personal records relevant to criminal history, taxes, entitlements, real-estate transactions, and scores of other matters—would not exist but for the government. The executive branch needs this information for all sorts of legitimate reasons, some related to crime prevention and many that are not. Government should not be prohibited from collecting and maintaining it.

Instead, regulation of program-driven Cloud searches must come from the political process.⁶⁶ Given Congress’ docility toward executive-branch surveillance proposals after 9/11, that suggestion may seem naïve. But legislatures are capable of action in this area, as the defunding of TIA and the revamping of the NSA’s metadata program illustrate.⁶⁷ Especially when, as is the case with many types of Cloud-based efforts, the program affects significant segments of the population—including members of the

⁶⁵ See Christopher Slobogin, *Panvasive Surveillance, Political Process Theory, and the Nondelegation Doctrine*, 102 GEO. L.J. 1721 (2014).

⁶⁶ *Id.* at 1745–58.

⁶⁷ Other examples are state statutes that limit the use of drone surveillance and federal statutes limiting access to various types of records. See Michael L. Smith, *Regulating Law Enforcement’s Use of Drones: The Need for State Legislation*, 52 HARV. J. LEGIS. 423, 427–32 (2015) (cataloguing state drone statutes); Murphy, *supra* note 5, at 546 (appendix detailing federal laws).

legislature and their most powerful constituents—some type of political oversight is not only possible but likely.

At the same time, it must be admitted that law enforcement and tough-on-crime lobbies are a forceful presence at both the federal and state levels and may be able to exert influence that the populace as a whole cannot. That is where the courts could come into play, in two ways. On rare occasions, courts might declare a particular data-collection scheme unconstitutional under the Fourth Amendment. However, given the Supreme Court's narrow definition of the word "search" for Fourth Amendment purposes and its high level of deference even to programs that it is willing to say involve searches (under what it calls its "special needs" jurisprudence⁶⁸), that outcome is not likely in the near future.

A second way courts might nudge legislatures and law enforcement agencies toward a balanced view—and one that would operate independently of the Fourth Amendment—is by applying the same "hard look" analysis they apply to programs created by other administrative agencies like the Environmental Protection Agency and the Food and Drug Administration.⁶⁹ While law enforcement departments have seldom been subject to the type of judicial monitoring to which other agencies routinely submit, that lack of oversight is likely a historical accident rather than a considered policy. The full argument for why courts are obligated to engage in such oversight will not be set out here.⁷⁰ For present purposes, it suffices to say that, where program-driven, panvasive operations are involved, a solid case can be made that the courts should treat police agencies the same way they treat other agencies that are engaged in creating rules governing the circumstances under which people may carry out innocent conduct.

That conclusion has several consequences. First, under accepted administrative law principles, no agency program that affects the rights and obligations of the citizenry may exist unless the agency can point to authorizing legislation that, ideally, sets out the harm to be prevented, the persons and activities likely to be affected, and the general means for preventing the harm. That would mean that before programs like New York City's Domain Awareness operation and the states' fusion centers can come into being, municipal, state, or federal legislatures would have to think through the types of information they can obtain and for what purpose. That

⁶⁸ For a description of this jurisprudence, see Slobogin, *Panvasive Surveillance*, *supra* note 65, at 1727–33.

⁶⁹ See, e.g., Patrick M. Garry, *Judicial Review and the "Hard Look" Doctrine*, 7 NEV. L.J. 151, 154–59 (2006) (discussing the Administrative Procedure Act and associated case law establishing the hard look doctrine).

⁷⁰ See Christopher Slobogin, *Policing as Administration*, 165 U. PA. L. REV. 91 (2016); see also Barry Friedman & Maria Ponomarenko, "Democratic Accountability and Policing," in the present Report.

requirement of legislative authorization, enforced by the courts, would ensure at least some democratic assessment of such programs and how they should operate.

The impact of administrative law principles would not end there, however. Standard practice dictates that, once authorized to set up a program, an agency must draft implementing rules, subject them to a notice-and-comment process (or something similar) that allows public input, and provide written rationales for the rules ultimately chosen—rules that are reviewable by a court to ensure they are consistent with the legislative delegation and that they are applied even-handedly, without irrational distinctions between groups or areas.⁷¹ This further injection of democratic input and judicial oversight would exert significantly more pressure on police departments to consider competing views when contemplating the creation of a data-collection scheme. Regulated through this type of public process, it is likely that TIA-like programs, fusion centers, and other panvasive practices would be significantly curtailed or implemented with more care.

The even-handedness requirement, designed to prevent biased data collection, is particularly important, so important that some have argued it should also be enforced through equal protection doctrine.⁷² It would call either for universal or random data collection (as suggested above in connection with profiles) or for proof that uneven information collection is justified statistically. For instance, this principle might demand that CCTV camera systems be established citywide or, alternatively, *everywhere* within the city that has similar reported crime rates. Metadata collection would be nationwide, random, or based on algorithms with high hit rates. And DNA database programs focused on arrestees, like the one authorized by the Supreme Court,⁷³ would be hard to justify without some proof that arrestees are significantly more likely to commit crimes than the general population.⁷⁴

One possible drawback to the political-process approach to program-driven Cloud searches is that its transparent nature will enable the bad guys to learn the ins-and-outs of the programs and how to avoid them. But this traditional law enforcement concern, which administrative procedure acts specifically recognize as legitimate,⁷⁵ is exaggerated in this setting. The

⁷¹ Slobogin, *Policing as Administration*, *supra* note 70, at 144–45.

⁷² Barry Friedman & Cynthia B. Stein, *Redefining What's "Reasonable": The Protections for Policing*, 84 GEO. WASH. L. REV. 281, 327–43 (2016).

⁷³ *Maryland v. King*, 567 U.S. 1301 (2012).

⁷⁴ Andrea Roth, *Maryland v. King and the Wonderful, Horrible DNA Revolution in Law Enforcement*, 11 OHIO ST. J. CRIM. L. 295, 308–09 (2013) (explaining the virtues of a universal DNA database).

⁷⁵ *See, e.g.*, 5 U.S.C. § 552(b)(7)(E) (2012).

primary aim of most panvasive actions is deterrence, which publicity can only enhance. Further, matters of specific implementation need not be revealed. For instance, if camera surveillance is meant to be covert, the fact and general area of such surveillance should be disclosed, but exact camera locations need not be. The types of records sought by fusion centers should be revealed, but the algorithms that might be used to analyze them could be viewed *in camera*. Ultimately, however, the primary response to the tip-off concern is that democratic accountability requires that the public be told not only what panvasive capacities police have but how those capacities will be used.

V. VOLUNTEER-DRIVEN CLOUD SEARCHES-FIDUCIARY OBLIGATIONS

All of the foregoing Cloud searches involve government-initiated investigations. The assumption throughout this paper has been that when the government decides to intrude, some justification is necessary. But what if a data-holder—a bank, a common carrier, or hospital—comes across information it thinks is indicative of criminal activity and wants to hand it over to the police? While the discussion thus far has suggested several reasons why government should not be able to demand information from a third party without justification, the situation is clearly different when the third party comes forward of its own accord.

Even so, it is important to recognize that not all volunteer-driven Cloud searches are alike. In the cases in which the Supreme Court first announced the third-party doctrine, the third party was a personal acquaintance of the defendant.⁷⁶ Establishing a rule that the government must ignore disclosures from such people denigrates their autonomous choice to make the disclosures, and could even be said to undermine their First Amendment right to speech. Recall, for instance, the tipster in the hypothetical involving John Slade. Whatever that person's motives and however that person acquired the information, the choice to divulge it deserves respect and should be considered a legitimate basis for government action if it has sufficient indications of reliability.

However, in the Court's later third-party cases, *Miller v. United States*⁷⁷ and *Smith v. Maryland*,⁷⁸ the third party was not a person but an institution, more specifically, a bank and a phone company. Historically, corporations have not been considered autonomous "persons" in most contexts and have also been accorded lesser First Amendment rights than

⁷⁶ See, e.g., *Lewis v. United States*, 385 U.S. 206 (1966); *Hoffa v. United States*, 385 U.S. 293 (1966).

⁷⁷ 425 U.S. 435 (1976).

⁷⁸ 442 U.S. 735 (1979).

natural beings.⁷⁹ More importantly, unlike human confidantes, these institutions can be said to owe either formal or quasi-formal fiduciary duties to their customers, because unlike the human third party, they are able to obtain personal facts solely because they purport to provide a particular service.⁸⁰ The most sympathetic example on point comes from the medical context, where a patient provides information to a treatment provider. Even the Supreme Court has balked at the notion that a hospital is entitled to ignore a patient's expectation of medical privacy for the purpose of catching criminals.⁸¹ Arguably, an analogous position is warranted with respect to banks and phone companies, to which we give information for the sole purpose of carrying out financial transactions or communicating.

Also important to recognize is that, when the third party is an institution, the degree to which information is "voluntarily" handed over to the government can vary greatly. In some cases, the government *commands* third parties to produce information about others, automatically and in the absence of a particularized court order. For instance, banks must report all deposits of \$10,000, regardless of circumstances.⁸² If this sort of command is justifiable, it should be so only if it comes from the legislature and is generally applicable (as is true in the deposit scenario). More commonly, the government exerts subtler pressures on third parties to produce information. Most obviously, some data brokers, although purportedly private and independent of the government, essentially see the government as their client,⁸³ and other companies, dependent on government largesse, may be especially eager to show they are helpful.⁸⁴ Unless defined narrowly, volunteer-driven Cloud searches might ultimately even undo

⁷⁹ The Court's decision in *Citizens United v. Fed. Elec. Comm'n*, 558 U.S. 310 (2010), focused on *political* speech rights of corporations, which are not implicated in this context. Further, corporations are still not considered "persons" for Fifth Amendment purposes, *see* *Hale v. Henkel*, 210 U.S. 43 (1906); and have very weak Fourth Amendment rights. *See* *United States v. Morton Salt Co.*, 338 U.S. 632, 652 (1950).

⁸⁰ *See* Kiel Brennan-Marquez, *Fourth Amendment Fiduciaries*, 84 *FORDHAM L. REV.* 611 (2015); SLOBOGIN, *supra* note 4, at 161 (arguing that recordholders have a fiduciary "duty of allegiance" to the subject of the record).

⁸¹ *Ferguson v. City of Charleston*, 532 U.S. 67 (2001).

⁸² 31 U.S.C. § 5313(a).

⁸³ Chris Jay Hoofnagle, *Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement*, 29 *N.C. J. INT'L L. & COM. REG.* 595, 617–18 (2004) (describing the FBI's "secret, classified contract" with Choicepoint).

⁸⁴ Avidan Y. Cover, *Corporate Avatars and the Erosion of the Populist Fourth Amendment*, 100 *IOWA L. REV.* 1441, 1445 (2015) ("[T]echnology corporations are not likely to challenge government surveillance requests, and even less likely to make effective arguments asserting their individual customers' rights, because of their government connections, the legal constraints on transparency and disclosure, and their immunity for complying with the government.").

efforts, like the recent NSA legislation, to keep as much data as possible out of government hands. That phenomenon is worrisome, because people should be able to trust that the private institutions on which they depend for the basics of life are not conduits to the government.

At the same time, it must be acknowledged that fiduciary obligations and concerns about corporate duplicity should not always trump speech rights and concerns about public safety. For instance, both the medical and legal professions recognize a duty to reveal information that would prevent a violent crime or forestall an ongoing one.⁸⁵ Explicitly applied to The Cloud, that norm would permit third-party institutions to disclose, and government to use, information about others that is likely to prevent a serious violent felony from taking place in the near future. Arguably, however, that norm should be the full extent to which the law bows to the volunteer notion where third-party institutions that are essential to living in the modern world are involved.⁸⁶

RECOMMENDATIONS

Databases are full of information that can enhance law enforcement's ability to detect and investigate crime and terrorism. Given the personal nature of much of this information, however, government should not be able to obtain, view, or use it at will. The following recommendations concerning law enforcement access to data arise out of the foregoing discussion.

- 1. If a policing agency seeks non-public records about an identified person, it should have to demonstrate suspicion of wrongdoing proportionate to the intrusion involved.** Whether or not courts modify current Fourth Amendment law to encompass such access, legislatures and agencies should require increasingly demanding justification requirements based on the nature of the data sought, the amount of data sought, or a combination thereof.

⁸⁵ MODEL RULES OF PROF'L CONDUCT Rule 1.6(b)(1) ("A lawyer may reveal information relating to the representation of a client to the extent the lawyer reasonably believes necessary ... to prevent reasonably certain death or substantial bodily harm."); FLA. STAT. ANN. § 394.4615(3)(a) ("When a patient has declared an intention to harm other persons," the therapist may release "sufficient information to provide adequate warning to the person threatened.").

⁸⁶ This is the rule Congress adopted in connection with communications. *See* 18 U.S.C. § 2702 (c) (prohibiting ISPs from disclosing communications to law enforcement except in emergencies involving death or serious physical injury and a few technical situations)

2. **If a law enforcement agency is instead accessing data for the purpose of executing a profile to identify suspects, it should ensure the profile produces the requisite proportionality-derived hit rate, avoids illegitimate discrimination, and uses an understandable algorithm.** Courts should evaluate these profiles, *in camera* if necessary, to ensure they are properly validated and do not rely on obviously biased risk factors. If the profile is used to identify suspects, police should not be able to choose whom among them will be subject to further investigation, but rather should be required to investigate all of those who meet the profile or, if that is not possible, a neutrally selected subset of that group.
3. **If policing agencies are relying on a crime rather than a suspect or a profile as the starting point of the investigation, the crime should be serious and the number of people investigated kept to the minimum dictated by the time and place of the crime.** At least when the investigation is extensive, judges should be involved in evaluating the need for and scope of such investigations.
4. **Collections of data needed by law enforcement should be maintained outside of government to the extent consistent with governing needs, but wherever maintained they should be authorized by specific legislation and administrative rules transparently and democratically arrived at.** Data-acquisition methods should be universal, random, or statistically justifiable. Courts should enforce these rules through either the administrative hard-look doctrine or equal protection analysis.
5. **Private institutions should be permitted to proffer to the government information about those to whom they owe a de facto fiduciary duty only when they have good reason to believe it would prevent an ongoing or future serious violent felony.** Courts should scrutinize any government incentives, financial or otherwise, that encourage the transfer of information that normally would be subject to the foregoing access and collection limitations.

These rules, accompanied by adequate accountability mechanisms that facilitate discovery of and sanctions for their breach,⁸⁷ would allow the government to take advantage of The Cloud's investigative potential while cabining the temptation to abuse it.

⁸⁷ Such mechanisms might include: (1) an auditing process indicating who accesses data, when, and for what purpose; (2) notice, either individualized (in the case of suspect-driven searches) or general (in other cases), detailing how Cloud access has occurred; (3) rules limiting data retention by the government or third parties; and (4) civil and criminal sanctions for wrongful collection or access. *See* SLOBOGIN, *supra* note 4, at 132-36.